

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: <b>C12N 9/50, 15/51, 15/57, 15/62, C12Q 1/37</b>		A2	(11) International Publication Number: <b>WO 96/36702</b> (43) International Publication Date: 21 November 1996 (21.11.96)
(21) International Application Number: <b>PCT/US96/06387</b> (22) International Filing Date: <b>9 May 1996 (09.05.96)</b>  (30) Priority Data: <b>08/440,409 12 May 1995 (12.05.95) US</b>		(81) Designated States: AL, AM, AU, AZ, BB, BG, BR, BY, CA, CN, CZ, EE, GE, HU, IS, JP, KG, KR, KZ, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(71) Applicant: SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033 (US).  (72) Inventors: DASMAHAPATRA, Bimalendu; 322 Grant Avenue, Nutley, NJ 07110 (US). MURRAY, Michael, G.; 8 Fawn Drive, Lebanon, NJ 08833 (US). RAMANATHAN, Lata; 32 Valley Way, West Orange, NJ 07052 (US). BUTKIEWICZ, Nancy, J.; 1000 Oak Lane, Plainfield, NJ 07060 (US).  (74) Agents: LUNN, Paul, G. et al.; Schering-Plough Corporation, Patent Dept. K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US).			
(54) Title: SOLUBLE, ACTIVE HEPATITIS C VIRUS PROTEASE			
(57) Abstract			
<p>Soluble HCV NS3 protease, including the NS3 protease fused to a solubilizing motif. A fusion of the NS3 and NS4 regions under conditions where they are not cleaved by the NS3 protease. Bacterially expressed soluble HCV NS3 protease. Host cells wherein at least 1 % of the cell's total protein is soluble HCV NS3 protease.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

SOLUBLE ACTIVE HEPATITIS C VIRUS PROTEASE

5

BACKGROUND OF THE INVENTION

10       Hepatitis C virus (HCV) is considered to be the major etiological agent of non-A non-B (NANB) hepatitis, chronic liver disease, and hepatocellular carcinoma (HCC) around the world. The viral infection accounts for greater than 90% of transfusion -associated hepatitis in U.S. and it is the predominant form of hepatitis in adults over 40 years of  
15       age. Almost all of the infections result in chronic hepatitis and nearly 20% develop liver cirrhosis.

The virus particle has not been identified due to the lack of an efficient *in vitro* replication system and the extremely low amount of  
20       HCV particles in infected liver tissues or blood. However, molecular cloning of the viral genome has been accomplished by isolating the messenger RNA (mRNA) from the serum of infected chimpanzees then cloned using recombinant methodologies. [Grakoui A. *et al. J. Virol.* 67: 1385 - 1395 (1993)] It is now known that HCV contains a positive strand  
25       RNA genome comprising approximately 9400 nucleotides, whose organization is similar to that of flaviviruses and pestiviruses . The genome of HCV, like that of flavi- and pestiviruses, encodes a single large polyprotein of about 3000 amino acids which undergoes proteolysis to form mature viral proteins in infected cells.

30       Cell-free translation of the viral polyprotein and cell culture expression studies have established that the HCV polyprotein is processed by cellular and viral proteases to produce the putative structural and nonstructural (NS) proteins. At least nine mature viral  
35       proteins are produced from the polyprotein by specific proteolysis. The order and nomenclature of the cleavage products are as follows: NH<sub>2</sub>-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH.(Fig 1). The three amino terminal putative structural proteins, C (capsid), E1, and E2 ( two

- 2 -

envelope glycoproteins), are believed to be cleaved by host signal peptidases of the endoplasmic reticulum(ER) . The host enzyme is also responsible for generating the amino terminus of NS2 . The proteolytic processing of the nonstructural proteins are carried out by the viral 5 proteases: NS2-3 and NS3, contained within the viral polyprotein. The NS2-3 protease catalyzes the cleavage between NS2 and NS3. It is a metalloprotease and requires both NS2 and the protease domain of NS3. The NS3 protease catalyzes the rest of the cleavages in the nonstructural 10 part of the polyprotein. The NS3 protein contains 631 amino acid residues and is comprised of two enzymatic domains: the protease domain contained within amino acid residues 1-181 and a helicase ATPase domain contained within the rest of the protein. It is not known if the 70 kD NS3 protein is cleaved further in infected cells to 15 separate the protease domain from the helicase domain, however, no cleavage has been observed in cell culture expression studies.

The NS3 protease is a member of the serine class of enzymes. It contains His, Asp, and Ser as the catalytic triad, Ser being the active site residue. Mutation of the Ser residue abolishes the cleavages at substrates 20 NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B. The cleavage between NS3 and NS4A is intramolecular, whereas the cleavages at NS 4A/4B, 4B/5A, 5A/5B sites occur in *trans* .

Experiments using transient expression of various forms of HCV 25 NS polyproteins in mammalian cells have established that the NS3 serine protease is necessary but not sufficient for efficient processing of all these cleavages. Like flaviviruses, the HCV NS3 protease also requires a cofactor to catalyze some of these cleavage reactions. In addition to the serine protease NS3, the NS4A protein is absolutely 30 required for the cleavage of the substrate at the 4B/5A site and increases the efficiency of cleavage of the substrate between 5A/5B, and possibly 4A/4B.

Because the HCV NS3 protease cleaves the non-structural HCV 35 proteins which are necessary for the HCV replication, the NS3 protease can be a target for the development of therapeutic agents against the HCV virus. The gene encoding the HCV NS3 protein has been cloned as disclosed in U.S. Patent No. 5,371,017, however, the protein has not

- 3 -

been produced in a soluble active form. If the HCV protease is to be useful as a target in a screen to discover therapeutic agents, the protease must be produced in a soluble active form. Thus, there is a need for a soluble active form of the HCV protease which can be produced in large 5 quantities to be used in high throughput screen to detect inhibitors of the protease and for structural studies.

#### SUMMARY OF THE INVENTION

10 The present invention fills this need by providing for a soluble, active NS3 protease. In one embodiment of the present invention, the soluble NS3 protease is contained within a fusion protein comprised of a HCV protease fused to a solubilizing motif.

15 The present invention further provides for a soluble fusion protein comprised of the catalytic domain of the NS3 protease, cofactor domain of cofactor NS4A and a solubilizing motif wherein the NS4A cofactor has been mutated so that the NS3 protease and NS4A cofactor are not cleaved by the catalytic activity of the NS3 protease.

20 The present invention further provides for an HCV NS3 protease having a polypeptide comprising three or more histidine residues fused to the protease. This enables rapid purification of the protease.

25 The present invention provides further for a soluble HCV NS3 protease selected from the group consisting of SEQ ID NO 3, SEQ ID NO 4, SEQ ID NO: 5 SEQ ID NO 7, SEQ ID NO 8, SEQ ID NO 9, SEQ ID NO 10 and SEQ ID NO 27.

30 The present invention further provides for isolated nucleic acids and vectors which encode the HCV proteases of the present invention, host cells transformed or transfected by said nucleic acids or vectors. Also claimed is a method for making a soluble HCV protease comprising culturing the transformed or transfected host cell under 35 conditions in which the nucleic acid or vector is expressed.

The present invention further provides for a host cell transformed or transfected with a nucleic acid or vector able to express

- 4 -

soluble HCV NS3 protease, wherein the soluble HCV NS3 protease which is expressed is at least 1%, 2%, 3%, 4%, 5% or more of the total protein expressed by the cell.

5

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 schematically depicts the HCV genome.

10 Figure 2 depicts the recombinant synthesis of plasmid pBJ1015.

Figure 3 depicts the recombinant synthesis of plasmid pTS56-9.

Figure 4 depicts the recombinant synthesis of plasmid pJB1006.

15

Figure 5 depicts the recombinant synthesis of plasmid pBJ1022.

Figure 6 depicts the recombinant synthesis of plasmid pNB(-V)182Δ4AHT.

20

Figure 7 depicts the recombinant synthesis of plasmid pT5His/HIV/183.

Figure 8 schematically depicts a high throughput assay for discovering HCV protease inhibitors using surface plasmon resonance technology.

25

DETAILED DESCRIPTION OF THE INVENTION

The teachings of all references cited are incorporated herein in their entirety by reference.

30

The present invention is the production of the HCV NS3 protease in a soluble form. The HCV NS3 protease must be in a soluble form to be used in a screen to detect compounds which inhibit the protease from cleaving its target substrate. We have discovered that if a peptide

35

containing a solubilizing motif is attached to either the NS3 protease, preferably to the carboxyl terminus, the NS3 protease becomes readily soluble.

- 5 -

The amino acid sequence of the NS3 protease catalytic domain is shown in SEQ ID NO: 1. Prior to the present invention the NS3 protease was not expressed in a cell in a soluble form in sufficient quantities for extraction and purification. Moreover, soluble HCV NS3 protease was not able to be produced in soluble form in bacteria. This is important because bacterial expression is the preferred method of expression of large quantities of HCV protease. Soluble HCV NS3 protease of the present invention can be produced in several ways. A solubilizing motif can be fused to the protein resulting in a soluble protein. A solubilizing motif is any chemical moiety bound to the HCV NS3 protease which results in the NS3 protease becoming soluble in a buffered solution. Examples of such solubilizing motifs are chains of amino acids having polar side chains, preferably positively charged amino acids. The chain of amino acids should be about 4 - 10 amino acid residues in length. The preferred amino acids are arginine and lysine. Another example of a solubilizing motif is an amphipathic moiety. The solubilizing motif can be fused to either the amino terminus or carboxy terminus of the NS3 protease. A sequence which has been successfully fused to the carboxyl terminus to produce soluble NS3 protease is -Arg - Lys - Lys - Lys - Arg - Arg- (SEQ ID NO: 2). This has been fused to the carboxyl end of the NS3 protease to produce the polypeptides of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 8 and SEQ ID NO: 27. Other examples of soluble HCV NS3 protease having a hydrophilic amino acid residue tail which were made are SEQ ID NO: 9, and SEQ ID NO: 10.

In another embodiment of the present invention, soluble HCV NS3 protease can also be produced which does not have a solubilizing motif as for example the proteases shown in SEQ ID NO: 1 and SEQ ID NO: 7. Preferably the NS3 protease will have a histidine tag fused to its amino acid terminus for use in purifying the protein on a nickel ( $Ni^{2+}$ ) coated resin. See SEQ ID NO: 5. In this embodiment the protease is produced as insoluble aggregates or as inclusion bodies in bacteria such as in *E. coli*.

35 The insoluble HCV NS3 protease is first extracted from the bacteria by homogenization or sonication of the bacteria. The aggregates containing the bacteria are then solubilized in a 5 M solution of

- 6 -

guanidine hydrochloride (GuHCl). The NS3 protease is then purified from high molecular weight aggregates by size exclusion chromatography, as for example by applying the solution to a SEPHACRYL S-300 size exclusion gel column. Fractions containing the

5 NS3 protease in 5 M GuCl are pooled and diluted to about 0.1 M GuHCl in a refolding buffer comprised of dithiothreitol and lauryl maltoside. The diluted solution is then applied to a reverse phase chromatography column and pools containing the NS3 protease collected. The pH of the protease fractions is then raised in a stepwise manner to about 7.4 so as

10 to produce properly refolded soluble, active NS3 protease.

It has also been discovered that the HCV NS3 protease is much more effective in cleaving the HCV non-structural proteins, if the co-factor NS4A protein is present (SEQ ID NO: 6). Accordingly, the present

15 invention is also comprised of a fusion of the NS4A cofactor domain protein with the NS3 protease, in particular the fusion of the NS3 protease and the NS4A cofactor wherein the NS4A is mutated such the NS3 protease and the NS4A cofactor is not cleaved by the NS3 protease. Examples of the fused NS3 and NS4A constructs are shown in SEQ ID

20 NOs, 7, 8, 9, 10 and 27.

DNA encoding the NS3 protease of this invention can be prepared by chemical synthesis using the known nucleic acid sequence [Ratner *et al.*, Nucleic Acids Res. 13:5007 (1985)] and

25 standard methods such as the phosphoramidite solid support method of Matteucci *et al.* [J Am. Chem. Soc. 103:3185 (1981)] or the method of Yoo *et al.* [J. Biol. Chem. 264:17078 (1989)]. See also Glick, Bernard R. and Pasternak, *Molecular Biotechnology* : pages 55 - 63, (ASM Press, Washington, D.C. 1994). The gene encoding the protease

30 can also be obtained using the plasmid disclosed in Grakoui, A., Wychowski, C., Lin, C., Feinstone, S. M., and Rice, C. M., Expression and Identification of Hepatitis C Virus polypeptide Cleavage Products, *J. Virol.* 67:1385-1395 (1993). Also, the nucleic acid encoding HCV protease can be isolated, amplified and cloned (from patients

35 infected with the HCV virus). Furthermore, the HCV genome has been disclosed in PCT WO 89/04669 and are available from the

American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, MD under ATCC accession no. 40394.

Of course, because of the degeneracy of the genetic code, there are many functionally equivalent nucleic acid sequences that can

5 encode mature human HCV protease as defined herein. Such functionally equivalent sequences, which can readily be prepared using known methods such as chemical synthesis, PCR employing modified primers and site-directed mutagenesis, are within the scope of this invention.

10 Various expression vectors can be used to express DNA encoding HCV NS3 protease. Conventional vectors used for expression of recombinant proteins used for expression of recombinant proteins in prokaryotic or eukaryotic cells may be used. Preferred vectors include

15 the pcD vectors described by Okayama *et al.*, *Mol. Cell. Bio.* 3: 280-289 (1983); and Takebe *et al.*, *Mol. Cell. Biol.* 8: 466-472 (1988). Other SV40-based mammalian expression vectors include those disclosed in Kaufman *et al.*, *Mol. Cell. Biol.* 2: 1304-1319 (1982) and U.S. Patent No. 4,675,285. These SV40-based vectors are particularly useful in COS7

20 monkey cells (ATCC No. CRL 1651), as well as in other mammalian cells such as mouse L cells and CHO cells.

Standard transfection methods can be used to produce eukaryotic cell lines which express large quantities of the polypeptide. Eukaryotic

25 cell lines include mammalian, yeast and insect cell lines. Exemplary mammalian cell lines include COS-7 cells, mouse L cells and Chinese Hamster Ovary (CHO) cells. See Sambrook *et al.*, *supra* and Ausubel *et al.*, *supra*.

30 As used herein, the term "transformed bacteria" means bacteria that have been genetically engineered to produce a mammalian protein. Such genetic engineering usually entails the introduction of an expression vector into a bacterium. The expression vector is capable of autonomous replication and protein expression relative to genes in the

35 bacterial genome. Construction of bacterial expression is well known in the art, provided the nucleotide sequence encoding a desired protein is known or otherwise available. For example, DeBoer in U.S. Pat. No.

4,551,433 discloses promoters for use in bacterial expression vectors; Goeddel *et al.* in U.S. Pat. No. 4,601,980 and Riggs, in U.S. Pat. No. 4,431,739 disclose the production of mammalian proteins by *E. coli* expression systems; and Riggs *supra*, Ferretti *et al.* *Proc. Natl. Acad. Sci.* 83:599 (1986), Sproat *et al.*, *Nucleic Acid Research* 13:2959 (1985) and Mullenbach *et al.*, *J. Biol. Chem.* 261:719 (1986) disclose how to construct synthetic genes for expression in bacteria. Many bacterial expression vectors are available commercially and through the American Type Culture Collection (ATCC), Rockville, Maryland.

10

Insertion of DNA encoding human HCV protease into a vector is easily accomplished when the termini of both the DNA and the vector comprise the same restriction site. If this is not the case, it may be necessary to modify the termini of the DNA and/or vector by 15 digesting back single-stranded DNA overhangs generated by restriction endonuclease cleavage to produce blunt ends, or to achieve the same result by filling in the single-stranded termini with an appropriate DNA polymerase. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the 20 termini. Such linkers may comprise specific oligonucleotide sequences that define desired restriction sites. The cleaved vector and the DNA fragments may also be modified if required by homopolymeric tailing.

Many *E. coli*-compatible expression vectors can be used to 25 produce soluble HCV NS3 protease of the present invention, including but not limited to vectors containing bacterial or bacteriophage promoters such as the *Tac*, *Lac*, *Trp*, *LacUV5*, *l P<sub>r</sub>* and *l P<sub>L</sub>* promoters. Preferably, a vector selected will have expression control sequences that permit regulation of the rate of HCV protease 30 expression. Then, HCV protease production can be regulated to avoid overproduction that could prove toxic to the host cells. Most preferred is a vector comprising, from 5' to 3' (upstream to downstream), a *Tac* promoter, a *lac I<sup>Q</sup>* repressor gene and DNA encoding mature human HCV protease. The vectors chosen for use 35 in this invention may also encode secretory leaders such as the *ompA* or protein A leader, as long as such leaders are cleaved during

- 9 -

post-translational processing to produce mature HCV protease or if the leaders are not cleaved, the leaders do not interfere with the enzymatic activity of the protease.

5        Fusion peptides will typically be made by either recombinant nucleic acid methods or by synthetic polypeptide methods. Techniques for nucleic acid manipulation and expression are described generally, e.g., in Sambrook, et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed.), vols. 1-3, Cold Spring Harbor Laboratory; and Ausubel, et al.  
10      (eds.) (1993) *Current Protocols in Molecular Biology*, Greene and Wiley, NY. Techniques for synthesis of polypeptides are described, e.g., in Merrifield (1963) *J. Amer. Chem. Soc.* 85:2149-2156; Merrifield (1986) *Science* 232: 341-347; and Stewart et al (1984)., "Solid Phase Peptide Synthesis" (2nd Edition), Pierce Chemical Co., Rockford, IL.; and  
15      Atherton, et al. (1989) *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, Oxford; and Grant (1992) *Synthetic Peptides: A User's Guide*, W.H. Freeman, NY.

20      The smaller peptides such as the NS4A cofactor and the substrates 5A/5B and 4B/5A can be synthesized by a suitable method such as by exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. The polypeptides are preferably prepared by solid phase peptide synthesis as described by Merrifield, *J. Am. Chem. Soc.* 85:2149 (1963). The synthesis is carried out  
25      with amino acids that are protected at the alpha-amino terminus. Trifunctional amino acids with labile side-chains are also protected with suitable groups to prevent undesired chemical reactions from occurring during the assembly of the polypeptides. The alpha-amino protecting group is selectively removed to allow subsequent reaction to take place  
30      at the amino-terminus. The conditions for the removal of the alpha-amino protecting group do not remove the side-chain protecting groups.

35      The alpha-amino protecting groups are those known to be useful in the art of stepwise polypeptide synthesis. Included are acyl type protecting groups (e.g., formyl, trifluoroacetyl, acetyl), aryl type protecting groups (e.g., biotinyl), aromatic urethane type protecting groups [e.g., benzyloxycarbonyl (Cbz), substituted benzyloxycarbonyl and 9-fluorenylmethoxy-carbonyl (Fmoc)],

- 10 -

aliphatic urethane protecting groups [e.g., t-butyloxycarbonyl (tBoc), isopropylloxycarbonyl, cyclohexyloxycarbonyl] and alkyl type protecting groups (e.g., benzyl, triphenylmethyl). The preferred protecting groups are tBoc and Fmoc, thus the peptides are said to be 5 synthesized by tBoc and Fmoc chemistry, respectively.

The side-chain protecting groups selected must remain intact during coupling and not be removed during the deprotection of the amino-terminus protecting group or during coupling 10 conditions. The side-chain protecting groups must also be removable upon the completion of synthesis, using reaction conditions that will not alter the finished polypeptide. In tBoc chemistry, the side-chain protecting groups for trifunctional amino acids are mostly benzyl based. In Fmoc chemistry, they are mostly 15 tert.-butyl or trityl based.

In tBoc chemistry, the preferred side-chain protecting groups are tosyl for Arg, cyclohexyl for Asp, 4-methylbenzyl (and acetamidomethyl) for Cys, benzyl for Glu, Ser and Thr, 20 benzyloxymethyl (and dinitrophenyl) for His, 2-Cl-benzyloxycarbonyl for Lys, formyl for Trp and 2-bromobenzyl for Tyr. In Fmoc chemistry, the preferred side-chain protecting groups are 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc) or 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) for Arg, trityl for 25 Asn, Cys, Gln and His, tert. butyl for Asp, Glu, Ser, Thr and Tyr, tBoc for Lys and Trp.

For the synthesis of phosphopeptides, either direct or post-assembly incorporation of the phosphate group is used. In the 30 direct incorporation strategy, the phosphate group on Ser, Thr or Tyr may be protected by methyl, benzyl or tert.butyl in Fmoc chemistry or by methyl, benzyl or phenyl in tBoc chemistry. Direct incorporation of phosphotyrosine without phosphate protection can also be used in Fmoc chemistry. In the post-assembly incorporation strategy, the 35 unprotected hydroxyl group of Ser, Thr or Tyr was derivatized on solid phase with di-tert.butyl-, dibenzyl- or dimethyl-N,N'-diisopropylphosphoramidite and then oxidized by tert.butylhydroperoxide.

Solid phase synthesis is usually carried out from the carboxyl-terminus by coupling the alpha-amino protected (side-chain protected) amino acid to a suitable solid support. An ester linkage is

5 formed when the attachment is made to a chloromethyl, chlortrityl or hydroxymethyl resin, and the resulting polypeptide will have a free carboxyl group at the C-terminus. Alternatively, when an amide resin such as benzhydrylamine or p-methylbenzhydrylamine resin (for tBoc chemistry) and Rink amide or PAL resin (for Fmoc

10 chemistry) is used, an amide bond is formed and the resulting polypeptide will have a carboxamide group at the C-terminus. These resins, whether polystyrene- or polyamide-based or polyethyleneglycol-grafted, with or without a handle or linker, with or without the first amino acid attached, are commercially available,

15 and their preparations have been described by Stewart et al (1984), "Solid Phase Peptide Synthesis" (2nd Edition), Pierce Chemical Co., Rockford, IL.; and Bayer & Rapp (1986) *Chem. Pept. Prot.* 3, 3; and Atherton, et al. (1989) *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, Oxford.

20

The C-terminal amino acid, protected at the side-chain if necessary and at the alpha-amino group, is attached to a hydroxymethyl resin using various activating agents including dicyclohexylcarbodiimide (DCC), N,N'-diisopropylcarbodiimide

25 DIPCDI) and carbonyldiimidazole (CDI). It can be attached to chloromethyl or chlorotriyl resin directly in its cesium tetramethylammonium salt form or in the presence of triethylamine (TEA) or diisopropylethylamine (DIEA). First amino acid attachment to an amide resin is the same as amide bond formation

30 during coupling reactions

Following the attachment to the resin support, the alpha-amino protecting group is removed using various reagents depending on the protecting chemistry (e.g., tBoc, Fmoc). The extent of Fmoc removal can be monitored at 300-320 nm or by a conductivity cell. After removal of the alpha-amino protecting group, the remaining protected amino acids are coupled stepwise in the required order to obtain the desired sequence.

Various activating agents can be used for the coupling reactions including DCC, DIPCDI, 2-chloro-1,3-dimethylimidium hexafluorophosphate (CIP), benzotriazol-1-yl-oxy-tris-

5 (dimethylamino)-phosphonium hexafluorophosphate (BOP) and its pyrrolidine analog (PyBOP), bromo-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBroP), O -(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and its tetrafluoroborate analog (TBTU) or its pyrrolidine analog (HBPyU),

10 O -(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and its tetrafluoroborate analog (TATU) or pyrrolidine analog (HAPyU). The most common catalytic additives used in coupling reactions include 4-dimethylaminopyridine (DMAP), 3-hydroxy-3,4-dihydro-4-oxo-1,2,3-

15 benzotriazine (HODhbt), N-hydroxybenzotriazole (HOBt) and 1-hydroxy-7-azabenzotriazole (HOAt). Each protected amino acid is used in excess (>2.0 equivalents), and the couplings are usually carried out in N-methylpyrrolidone (NMP) or in DMF, CH<sub>2</sub>Cl<sub>2</sub> or mixtures thereof. The extent of completion of the coupling reaction

20 can be monitored at each stage, e.g., by the ninhydrin reaction as described by Kaiser *et al.*, *Anal. Biochem.* 34:595 (1970). In cases where incomplete coupling is found, the coupling reaction is extended and repeated and may have chaotropic salts added. The coupling reactions can be performed automatically with

25 commercially available instruments such as ABI model 430A, 431A and 433A peptide synthesizers.

After the entire assembly of the desired polypeptide, the polypeptide-resin is cleaved with a reagent with proper scavengers.

30 The Fmoc peptides are usually cleaved and deprotected by TFA with scavengers (e.g., H<sub>2</sub>O, ethanedithiol, phenol and thioanisole). The tBoc peptides are usually cleaved and deprotected with liquid HF for 1-2 hours at -5 to 0°C, which cleaves the polypeptide from the resin and removes most of the side-chain protecting groups. Scavengers

35 such as anisole, dimethylsulfide and p-thiocresol are usually used with the liquid HF to prevent cations formed during the cleavage from alkylating and acylating the amino acid residues present in the polypeptide. The formyl group of Trp and dinitrophenyl group of His

need to be removed, respectively, by piperidine and thiophenol in DMF prior to the HF cleavage. The acetamidomethyl group of Cys can be removed by mercury(II) acetate and alternatively by iodine, thallium (III) trifluoroacetate or silver tetrafluoroborate which

5 simultaneously oxidize cysteine to cystine. Other strong acids used for tBoc peptide cleavage and deprotection include trifluoromethanesulfonic acid (TFMSA) and trimethylsilyltrifluoroacetate (TMSOTf).

10 Recombinant DNA methodology can also be used to prepare the polypeptides. The known genetic code, tailored if desired with known preferred codons for more efficient expression in a given host organism, can be used to synthesize oligonucleotides encoding the desired amino acid sequences. The phosphoramidite  
15 solid support method of Matteucci *et al.*, *J. Am. Chem. Soc.* 103:3185 (1981) or other known methods can be used for such syntheses. The resulting oligonucleotides can be inserted into an appropriate vector and expressed in a compatible host organism.

20 The polypeptides of the invention can be purified using HPLC, gel filtration, ion exchange and partition chromatography, countercurrent distribution or other well known methods. In a preferred embodiment of the present invention the NS3 fusion proteins also contain a histidine tag which facilitates purification using a Ni<sup>+</sup> column as is  
25 illustrated below.

One can use the NS3 protease, the NS4 cofactor and the peptide substrates, either 4B/5A or 5A/5B, to develop high throughput assays. These can be used to screen for compounds which inhibit proteolytic  
30 activity of the protease. This is carried out by developing techniques for determining whether or not a compound will inhibit the NS3 protease from cleaving the viral substrates. Examples of such synthetic substrates are SEQ ID NOs 16, 17, 18, 19, 20 and 21. If the substrates are not cleaved, the virus cannot replicate. One example of such a high throughput  
35 assay is the scintillation proximity assay (SPA). SPA technology involves the use of beads coated with scintillant. Bound to the beads are acceptor molecules such as antibodies, receptors or enzyme substrates which interact with ligands or enzymes in a reversible manner.

For a typical protease assay the substrate peptide is biotinylated at one end and the other end is radiolabelled with low energy emitters such as  $^{125}\text{I}$  or  $^3\text{H}$ . The labeled substrate is then incubated with the 5 enzyme. Avidin coated SPA beads are then added which bind to the biotin. When the substrate peptide is cleaved by the protease, the radioactive emitter is no longer in proximity to the scintillant bead and no light emission takes place. Inhibitors of the protease will leave the substrate intact and can be identified by the resulting light emission 10 which takes place in their presence.

Another type of protease assay, utilizes the phenomenon of surface plasmon resonance (SPR). A novel, high throughput enzymatic assay utilizing surface plasmon resonance technology has been 15 successfully developed. Using this assay, and a dedicated BIAcore™ instrument, at least 1000 samples per week can be screened for either their enzymatic activity or their inhibitory effects toward the enzymatic activity, in a 96 well plate format. This methodology is readily adaptable to any enzyme-substrate reaction. The advantage of this assay over the 20 SPA assay is that it does not require a radiolabeled peptide substrate.

The following examples are included to illustrate the present invention but not to limit it.

25

#### Examples 1

##### Production of HCV NS3 Protease

30 A. Plasmid constructions.

Several plasmids were designed and constructed using standard recombinant DNA techniques (Sambrook, Fritsch & Maniatis) to express the HCV protease in *E. coli* (Fig 2-7). All HCV specific sequences 35 originated from the parental plasmid pBRTM/HCV 1-3011 (Grakoui *et al.* 1993). To express the N-terminal 183 amino acid versions of the protease, a stop codon was inserted into the HCV genome using synthetic oligonucleotides (Fig. 3). The plasmids designed to express the

- 15 -

N-terminal 246 amino acid residues were generated by the natural NcoI restriction site at the C-terminus.

i) Construction of the plasmid pBJ1015 (Figure 2)

5

The plasmid pBRTM/HCV 1-3011 containing the entire HCV genome (Grakoui A., et al., *J. Virol.* 67: 1385-1395) was digested with the restriction enzymes Sca I and Hpa I and the 7138 bp (base pair) DNA fragment was isolated and cloned to the Sma I site of pSP72 (Promega) to

10 produce the plasmid pRJ201. The plasmid pRJ 201 was digested with Msc I and the 2106 bp Msc I fragment was isolated and cloned into the Sma I site of the plasmid pBD7. The resulting plasmid pMBM48 was digested with Kas I and Nco I, and the 734 bp DNA fragment after blunt ending with Klenow polymerase was isolated and cloned into Nco I  
15 digested, klenow polymerase treated pTrc HIS B seq expression plasmid (Invitrogen). The ligation regenerated a Nco I site at the 5' end and Nsi I site at the 3' end of HCV sequence. The plasmid pTHB HCV NS3 was then digested with Nco I and Nsi I, and treated with klenow polymerase and T4 DNA polymerase, to produce a blunt ended 738 bp DNA  
20 fragment which was isolated and cloned into Asp I cut, klenow polymerase treated expression plasmid pQE30 (HIV). The resulting plasmid pBJ 1015 expresses HCV NS3 (246 amino acids) protease.

(ii) Construction of the plasmid pTS 56-9 with a stop codon after

25 amino acid 183 (Figure 3)

The plasmid pTHB HCV NS3 was digested with Nco I, treated with klenow polymerase, then digested with Bst Y I; and the DNA fragment containing HCV sequence was isolated and cloned into Sma I  
30 and Bgl II digested pSP72. The resulting plasmid pTS 49-27 was then digested with Bgl II and Hpa I and ligated with a double stranded oligonucleotide:

GA TCA CCG GTC TAG ATCT

T GGC CAG ATC TAGA (SEQ ID NO 11) to produce pTS 56-9.

35 Thus, a stop codon was placed directly at the end of DNA encoding the protease catalytic domain of the NS3 protein. This enabled the HCV protease to be expressed independently from the helicase domain of the NS3 protein.

(iii) Construction of the plasmid pJB 1006 Fused with a peptide of positively charged amino acids at the carboxy terminus of NS3 183 (Figure 4).

5

The plasmid pTS 56-9 was digested with Sph I and Bgl II and the DNA fragment containing HCV sequence was isolated and cloned into a Sph I, Bgl II cut pSP72. The resulting plasmid pJB 1002 digested with Age I and HpaI and ligated to a double stranded oligonucleotide,

10

CCG GTC CGG AAG AAA AAG AGA CGC TAG C  
AG GCC TTC TTT TTC TCT GCG ATC G

(SEQ ID NO 12), to construct pJB 1006. This fused the hydrophilic, solubilizing motif onto the NS3 protease.

15

(iv) Construction of the plasmid pBJ 1022 expressing His-NS3(183)-HT in E.coli (Figure 5)

20

The plasmid pJB 1006 was digested with NgoM I and Nhe I and the 216 bp DNA fragment was isolated and cloned into Ngo M I, Nhe I cut pBJ 1015 to construct plasmid pBJ 1019. The plasmid pBJ 1019 was digested with Nar I and Pvu II, and treated with Klenow polymerase to fill in 5' ends of Nar I fragments. The expression plasmid pQE31 (Invitrogen) was digested with BamH I, blunt ended with Klenow polymerase. The 717 bp

25

Nar I- Pvu II DNA fragment was isolated and ligated to the 2787 bp BamH I/Klenowed -Msc I (Bal I) fragment of the expression plasmid pQE31 (Invitrogen). The recombinant plasmid, pBJ 1022, obtained after transformation into E.coli expresses His NS3(2-183)-HT which does not contain any HIV protease cleavage site sequence. The plasmid also

30

contains a large deletion in the CAT (Chloramphenicol Acetyl Transferase) gene.

(v) Construction of the plasmid pNB(-V)182-Δ4A HT (Figure 6)

35

The plasmid pMBM 48 was digested with Eag I and Xho I, treated with Klenow polymerase and the 320 bp DNA fragment was isolated and cloned into BamH I cut , blunt ended pSP 72 to construct the plasmid pJB1004. The 320 bp fragment encodes 7 amino acid from carboxy

terminal of NS3(631), all of NS4A, and the amino terminal 46 amino acid of NS4B. The recombinant plasmid pJB1004 was digested with Eag I and Cel 2, blunt ended with Klenow polymerase. The 220 bp DNA fragment was isolated and cloned into the expression plasmid pQE30

5 which was digested with BamH I and blunt ended with Klenow polymerase prior to ligation. The resulting plasmid pJB 1011 was digested with NgoM I and Hind III and ligated to a double stranded oligonucleotide ,

10 CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAA TTC  
GT TAA TAT GGA CTG TCC CTC CAA GAG ATG GTC CTT AAG

GAT GAG ATG GAA GAG TGC CGG AAG AAA AAG AGA CGC A  
CTA CTC TAC CTT CTC ACC CCC TTC TTT TTC TCT GCG TTC GA

15 (SEQ ID NO 13)

to construct the plasmid pNB 4A HT. The plasmid pNB 4AHT was digested with Msl I and Xba I. The 1218 bp DNA fragment was isolated and cloned into Age I cut, klenow polymerase treated, Xba I cut vector

20 DNA of pBJ 1019. The ligation results in a substitution of the 183rd amino acid residue valine by a glycine residue in NS3, and a deletion of amino terminal three amino acid residues of NS4A at the junction. The recombinant plasmid pNB182Δ4A HT comprising NS3(182aa)-G-NS4A(4-54 amino acid) does not contain NS3/NS4A cleavage site

25 sequence at the junction and is not cleaved by the autocatalytic activity of NS3. Finally the plasmid pNB182Δ4A HT (SEQ ID NO 8) was digested with Stu I and Nhe I, the 803 bp DNA fragment was isolated and cloned into Stu I and Nhe I cut plasmid pBJ 1022. The resulting plasmid pNB(-V)182-Δ4A HT contains a deletion of the HIV sequence from the amino

30 terminus end of the NS3 sequence and in the CAT gene (SEQ ID NO 27).

(vi) Construction of the plasmid pT5 His HIV-NS3 (Figure 7)

The plasmid pTS56-9 was digested with Bgl II, and treated with 35 Klenow polymerase to fill in 5' ends. The plasmid was then digested with NgoM I and the blunt ended Bgl II/NgoMI fragment containing the NS3 sequence was isolated and ligated to the Sal I, Klenow treated

- 18 -

Ngo MI cut and Sal I klenowled pBJ 1015. The resulting plasmid is designated pT5His HIV 183.

Example 2

5

Purification of HCV NS3 Protease having a Solubilizing Motif

Purification of His182HT (SEQ ID NO 4) and  
His (-V)182Δ4AHT (SEQ ID NO 8)

10

The recombinant plasmids pBJ1022 and pNB(-V)182Δ4A were used to transform separate cultures of *E. coli* strain M15 [pREP4] (Qiagen), which over-expresses the *lac* repressor, according to methods recommended by the manufacturer. M15 [pREP4] bacteria harboring

15 recombinant plasmids were grown overnight in broth containing 20g/L bactotryptone, 10g/L bacto-yeast extract, 5g/L NaCl and supplemented with 100 $\mu$ g/ml ampicillin and 25 $\mu$ g/ml kanamycin. Cultures were diluted down to O.D.600 of 0.1, then grown at 30°C to O.D.600 of 0.6 to 0.8, after which IPTG was added to a final concentration of 1mM. At

20 post-induction 2 to 3 hours, the cells were harvested by pelleting, and the cell pellets were washed with 100mM Tris, pH 7.5. Cell lysates were prepared as follows: to each ml equivalent of pelleted fermentation broth was added 50 $\mu$ l sonication buffer (50mM sodium phosphate, pH 7.8, 0.3M NaCl) with 1mg/ml lysozyme; cell suspension was placed on

25 ice for 30 min. Suspension was then brought to a final concentration of 0.2% Tween-20, 10mM dithiothreitol (DTT), and sonicated until cell breakage was complete. Insoluble material was pelleted at 12,000 x g in a microcentrifuge for 15 minutes, the soluble portion was removed to a separate tube and the soluble lysate was then brought to a final

30 concentration of 10% glycerol. Soluble lysates from cells expressing the plasmids produce strongly immunoreactive bands of the predicted molecular weight. Soluble lysates prepared for Ni<sup>2+</sup> column purification were prepared with 10mM  $\beta$ -mercaptoethanol (BME) instead of DTT. Lysates were stored at -80°C.

35

Purification using Ni<sup>2+</sup>-Nitrosyl acetic acid (NTA) agarose (OIAGEN)

The proteins were then purified by placing the extracted lysate on an NTA agarose column. NTA agarose column chromatography was used because the histidine tag which was fused to the N-terminus of the proteases readily binds to the nickel column. This produces a powerful affinity chromatographic technique for rapidly purifying the soluble protease. The column chromatography was performed in a batch mode.

5      The Ni<sup>2+</sup> NTA resin (3ml) was washed twice with 50 ml of Buffer A (50mM sodium phosphate pH 7.8 containing 10% glycerol, 0.2% Tween-20, 10mM BME). The lysate obtained from a 250 ml fermentation (12.5 ml) was incubated with the resin for one hour at 4°C. The flow through was collected by centrifugation. The resin was packed into a 1.0 x 4 cm

10     column and washed with buffer A until the baseline was reached. The bound protein was then eluted with a 20 ml gradient of imidazole (0-0.5M) in buffer A. Eluted fractions were evaluated by SDS-PAGE and western blot analysis using a rabbit polyclonal antibody to His-HIV 183. The amount of soluble, active, HCV protease recovered was equal to

15     about 5% of the total protein expressed by the cells as determined by the Bradford assay, U.S. Patent No. 4,023,933.

20    

Purification using POROS metal-chelate affinity column

25     In an alternative method to purify the proteins the lysate containing the proteins were applied to a POROS metal-chelate affinity column. Perfusion chromatography was performed on a POROS MC metal chelate column (4.6 x 50mm, 1.7 ml) precharged with Ni<sup>2+</sup>. The sample was applied at 10 ml/min and the column was washed with buffer A.

30     The column was step eluted with ten column volumes of buffer A containing 25 mM imidazole. The column was further eluted with a 25 column volume gradient of 25-250 mM imidazole in buffer A. All eluted fractions were evaluated by SDS-PAGE and western blot analysis using rabbit polyclonal antibody. The amount of soluble, active, HCV protease recovered was equal to about 5% of the total protein expressed by the cells as determined by the Bradford assay.

35

**Example 3**

5

**Peptide Synthesis of the 5A/5B and 4B/5A Substrates**

The peptides 5A/5B and 4B/5A substrates (SEQ ID NOs 16, 18, 19, 20 and 21) were synthesized using Fmoc chemistry on an ABI model 431A peptide synthesizer. The manufacturer recommended FastMoc™ activation strategy (HBTU/HOBt) was used for the synthesis of 4A activator peptide. A more powerful activator, HATU with or without the additive HOAt were employed to assemble 5A/5B substrate peptides on a preloaded Wang resin. The peptides were cleaved off the resin and deprotected by standard TFA cleavage protocol. The peptides were purified on reverse phase HPLC and confirmed by mass spectrometric analysis.

**Example 4**

20

**HPLC-assay using a synthetic 5A/5B peptide substrate**

To test the proteolytic activity of the HCV NS3 protease the DTEDVVCC SMSYTWTGK (SEQ ID NO 16) and soluble HCV NS3 (SEQ ID NO 27) were placed together in an assay buffer. The assay buffer was 25 50mM sodium phosphate pH 7.8, containing 15% glycerol, 10mM DTT, 0.2% Tween20 and 200 mM NaCl). The protease activity of SEQ ID NO 27 cleaved the substrate into two byproduct peptides, namely 5A and 5B. The substrate and two byproduct peptides were separated on a reversed-phase HPLC column. (Dynamax, 4.6 x 250 mm) with a pore size of 300Å 30 and a particle size of 5µm. The column was equilibrated with 0.1%TFA (Solvent A) at a flow rate of 1 ml per minute. The substrate and the product peptide standards were applied to the column equilibrated in A. Elution was performed with a acetonitrile gradient (Solvent B=100% acetonitrile in A). Two gradients were used for elution (5% to 70% B in 35 50 minutes followed by 70% to 100% B in 10 minutes).

In another experiment, partially purified SEQ ID NO 27 or vector control was incubated with 100µM of substrate for 3, 7 and 24 hours at

- 21 -

30°C. The reaction mixture was quenched by the addition of TFA to 0.01% and applied to the reversed-phase HPLC column. The fractions from each run were evaluated by mass spectrometry and sequencing.

5

Example 5

Analysis of NS3 Protease Activity By *In Vitro* Translation Assay

10

To detect HCV NS3 protease activity in *trans*, we have expressed a 40 kD protein containing the NS5A/5B cleavage site in cell-free translation system and used that as the substrate for the enzyme. The substrate protein produces two protein products of apparent molecular weight 12.5 kD (NS 5A') and 27 kD (NS5B') upon cleavage by the HCV NS3 protease.

The plasmid pTS102 encoding the substrate 5A/5B was linearized by digestion with EcoR I and was transcribed using T7 RNA polymerase 20 *in vitro*. The RNA was translated in presence of <sup>35</sup>S methionine in rabbit reticulocyte lysates according to the manufacturer's (Promega ) protocol to produce HCV specific protein. In a 20  $\mu$ l total reaction mixture containing 10mM Tris, pH 7.5, 1mM DTT, 0.5mM EDTA, and 10% glycerol was placed 2 to 8  $\mu$ l of <sup>35</sup>S methionine-labeled translated 25 5A/5B substrate. The reaction was started with the addition of 10 $\mu$ l of HCV NS3 protease in solubilization buffer (50mM Na Phosphate, pH 7.8, 0.3M NaCl, 0.2% Tween 20, 10 mM DTT or BME, 10% glycerol), and incubated at 30°C for the specified time. Reactions were stopped by adding an equal volume of 2X Laemmli sample buffer (Enprotech Inc.) 30 and heating at 100°C for 3 minutes. Reaction products were separated by SDS PAGE electrophoresis; gels were fixed, dried and subjected to autoradiography.

The *in vitro* translated substrate was used to assay the HCV NS3 35 proteases expressed by *E. coli* harboring plasmids pBJ1022 and pNB(-V)182 $\Delta$ 4A (SEQ ID NOs 4 and 27). In a two hour assay incubated at 30°C, pBJ1022 crude soluble lysate at 3, 6, and 10 $\mu$ l, was able to cleave 5A/5B substrate in a dose responsive manner, producing the expected

- 22 -

cleaved products: 5A (12.5 kD) and 5B (27 kD) as shown by SDS PAGE analysis. Corresponding vector control lysate did not show any cleavage activity over background. The crude soluble lysate derived from pNB182Δ4A was much more active in this assay. After only 30 minutes 5 incubation, the 5A and 5B cleavage products were detected using as little as 0.125μl cell lysate, with increasing amounts of lysate showing increased cleavage, reaching a maximum at 1μl.

10 We performed a time course study of the NS3 Protease activity of pNB182Δ4A in an *in vitro* translation assay for further characterization of the activity. At 30°C, in a reaction containing the translated 5A/5B substrate plus pNB182Δ4A soluble lysate at 1μl per 20μl reaction volume, the 5A and 5B cleavage products appeared beginning at 1 minute, and increased with time at 2.5, 5, 10, and 20 minutes.

15 Since we were able to demonstrate HCV NS3 Protease activity using crude cell lysates of pBJ1022 and pNB182Δ4A, we wanted to at least partially purify the expressed proteins in an effort to remove bacterial proteases from these preparations. For this purpose, affinity column 20 chromatography using Ni<sup>2+</sup> bound ligands was found to be effective, binding the histidine tag at the amino terminal ends of the expressed proteins, and subsequently releasing the bound proteins by imidazole elution. The imidazole-eluted fractions resulting from the purification of pNB182Δ4A on a Ni-NTA column were tested for activity in the *in* 25 *vitro* translation assay. The resultant fractions were all able to cleave the translated 5A/5B Substrate, producing the expected 5A and 5B products. Background bacterial protease activity was not detected in these eluted fractions.

30 As was described above, pBJ1022 was purified by another method of Ni<sup>2+</sup> chelate chromatography, using POROS Ni<sup>2+</sup> chelate resin and perfusion chromatography. Imidazole-eluted fractions which were positive for immunoreactivity with antibody to NS3 183 were tested for HCV protease activity by *in vitro* translation assay. In order to optimize 35 detection of activity in this assay for HCV protease, reactions were supplemented with a truncated peptide derived from the NS4A cofactor which has been shown to enhance cleavage at the 5A/5B site by NS3 protease. The cofactor was supplied as a synthetic peptide containing

amino acids 22 to 54 of NS4A (strain HCV-BK) at a final concentration of 1 $\mu$ M. All fractions tested were active in this translation assay.

Example 6

5

ENHANCEMENT BY 4A PEPTIDES

NS4A is able to enhance the NS3 serine protease activity at

10 NS5A/5B site in mammalian cells that transiently coexpress NS3, NS4A, and the various HCV non-structural polyprotein containing downstream cleavage sites. We have studied this enhancement activity in a well defined cell-free biochemical assay, using the partially purified *E.coli*-expressed pBJ1022 as a source of NS3 protease, and synthetic

15 peptides containing various truncations of NS4A. In our first experiment we used a crude cell lysate of pBJ1022 as the enzyme and NS4A synthetic peptide truncated 33 mer from amino acid 22 to amino acid 54, the carboxy-terminal *in vitro* translation cleavage reaction. The C-terminal 33 amino acid peptide of NS4A was able to enhance the

20 activity of the NS3 catalytic domain in a dose dependent manner from 0.01  $\mu$ M to 1.0  $\mu$ M peptide, producing the expected products of 5A (12.5kD), and 5B (27kD) from the 40kD translated 5A/5B substrate. Without the 4A peptide a relatively low cleavage activity by the protease alone was observed at the short incubation time of 30 minutes. The 4A

25 peptide itself or with the combination of crude lysate produced from cells harboring the vector plasmid did not cleave the substrate.

To further characterize NS4A enhancement activity additional truncations were made to the NS4A sequence. Truncated peptides were

30 evaluated for their activity in the *in vitro* translation assay using Ni<sup>2+</sup> chelate column-purified pBJ1022 (NS3 catalytic domain). We observed that in addition to the C-terminal 33 amino acid peptide, a 18 amino acid peptide containing the NS4A sequence from amino acid 19 through 36 was able to enhance the NS3 mediated cleavage activity. Other peptides,

35 including the N-terminal 21 amino acid, and two shorter truncations from the carboxyl terminal end, a 22mer and a 15mer, were found to have no effect; also a heterologous peptide of 18 amino acid also had no enhancement activity.

### Discussion

The experiments described in this report clearly demonstrate that

5 bacterially expressed HCV protease catalyzes cleavage of i) HCV polyproteins and ii) synthetic peptide substrates in *trans* biochemical assay. The processing activity of NS3 catalytic domain is enhanced by NS4A and its derivatives. The activity of the fusion protein containing

10 the NS3 catalytic domain and NS4A is much superior to that of the NS3 catalytic domain alone.

Hydrophobicity analysis of the catalytic domain of the NS3 protease reveals that the protein is very hydrophobic and also it contains seven cysteine residues. To neutralize hydrophobicity and thus to

15 improve solubility we have added six positively charged amino acid residues as a solubilizing motif. The addition of a solubilizing motif appears to improve the solubility without affecting the enzymatic activity.

20 We have also shown that the HCV NS4A from Japanese BK strain has enhanced the HCV-H NS3 mediated cleavage at 5A/5B site. This suggests that essential elements of recognition may be conserved among various strains of HCV.

25 It is clear from above experimental results that attachment of hydrophilic tail (solubilizing motif/water attracting structures) at the carboxy terminal end of histidine fused NS3 catalytic domain improved expression of soluble protein in *E.coli*. In these experiments six residues of positively charged amino acids are attached at the carboxy terminal

30 end of the protein. Another example of a solubilizing motif is an amphipathic helix tail (peptides having charged and hydrophobic amino acid residues to form both charged and hydrophobic faces) which is fused to the HCV NS3 protease. Addition of an amphipathic helix at the carboxy terminus of such fusion proteins will be an alternative way

35 to achieve improvement of solubility without affecting the enzymatic activity of the protease.

- 25 -

The hydrophilic tail used in these experiments consists of six amino acids. The sequence and length of the hydrophilic amino acids can be varied to achieve optimal expression of soluble protein. Therefore size of the solubilizing motif and nature of charged residues 5 may effect the expression of soluble NS3 in *E.coli*.

Position of these water attracting structures/motifs at both ends, at one end (amino terminal or carboxy terminal), or insertion within the NS3 catalytic domain and NS3 (catalytic domain)-4A fusion protein, 10 may improve solubility of the protein without affecting the activity.

Based on sequence homology to the members of trypsin superfamily and the protease of other members of the flaviviruses, it is predicted that the amino terminal 181 amino acid of NS3 is the catalytic 15 domain of HCV NS3 protease. Recently it has also been shown that a protein of 169 amino acid containing a 10 amino acid deletion from the amino terminus and 2 amino acid from carboxy terminal of the catalytic domain retains full enzymatic activity. The model we have developed predicts that a protein of 154 amino acids containing a deletion of 26 20 amino acid from amino terminal and a deletion of 2 amino acid from the carboxyl terminus would retain full enzymatic activity for the 5A/5B substrate.

Analysis of the amino acid sequence of the catalytic domain of 25 NS3 protease reveals that the protein contains seven cysteine residues, an odd number, which may cause aggregation. Mutation of one cysteine residue (located on the surface of the protein molecule and not involved in the activity) may improve solubility of the protein without affecting the protease activity.

30 Using the cell free biochemical assay we have demonstrated that the synthetic peptide containing 18 amino acid of HCV NS4A protein is sufficient to enhance the cleavage at NS5A/5B site mediated by the catalytic domain of NS3.

35

Example 7Refolding of Insoluble HCV NS3 Protease

5

The present example describes a novel process for the refolding of HCV NS3 protease which does not have a solubilizing motif from an *E.coli* inclusion body pellet. This procedure can be used to generate purified enzyme for activity assays and structural studies.

10

Extraction and Purification of His-HIV 183 from the *E.coli* inclusion body pellet

15

*E. coli* cells harboring the plasmid for HisHIV183 was used to transform a culture of *E. coli* strain M15 [pREP4] (Qiagen), which over-expresses the *lac* repressor, according to methods recommended by commercial source. M15 [pREP4] bacteria harboring recombinant plasmids were grown overnight in 20-10-5 broth supplemented with

20

100 $\mu$ g/ml ampicillin and 25 $\mu$ g/ml kanamycin. Cultures were diluted to O.D.600 of 0.1, then grown at 37°C to O.D.600 of 0.6 to 0.8, after which IPTG was added to a final concentration of 1mM. At post-induction 2 to 3 hours, the cells were harvested by pelleting, and the cell pellets were washed with 100mM Tris, pH 7.5. were pelleted by centrifugation. The

25

cell pellet was resuspended in 10 ml of 0.1M Tris-HCl, 5mM EDTA, pH 8.0 (Buffer A) for each gm wet weight of pellet. The pellet was homogenized and resuspended using a Dounce homogenizer. The suspension was clarified by centrifugation at 20,000 x g for 30 minutes at 4°C. The pellet was sequentially washed with the following five buffers:

30

1. Buffer A
2. 1.0M sodium chloride (NaCl) in buffer A
3. 1.0% Triton X-100 in buffer A
- 35 4. Buffer A
5. 1.0 M Guanidine HCl ( GuHCl) in buffer A.

The washed pellet was solubilized with 5M GuHCl, 1% beta mercaptoethanol in buffer A (3 ml per gm wet wt. of pellet) using a Dounce homogenizer and centrifuged at 100,000 x g for 30 minutes at 4°C. Purification of denatured HisHTV183 from high 5 molecular weight aggregates was accomplished by size exclusion on a SEPHACRYL S-300 gel filtration column.

In particular, an 8 ml sample of the 5.0M GuHCl *E. coli* extract was applied to a 160 ml Pharmacia S-300 column (1.6 x 100 cm) at a flow 10 rate of 1.0 ml/min. The column buffer was comprised of 5.0 M GuHCl, 0.1 M Tris-HCl, pH 8.0, and 5.0 mM EDTA. The fraction size was 5.0 ml. Appropriate fractions were pooled based on the results of SDS-PAGE, as well as N-terminal sequence analysis of the protein transferred to a Pro-Blot.

15 Detergent-assisted refolding of HCV-protease

The protein was concentrated by ultrafiltration using a 43 mm Amicon YM10 membrane to 1.0 mg per ml in 5M GuHCl, 0.1M Tris-HCl 20 pH 8.0, 1.0 mM EDTA, 1.0% beta-mercaptopethanol. It was then diluted 50-fold to 0.1M GuHCl in refolding buffer (100 mM sodium phosphate pH 8.0, 10mM DTT, 0.1% lauryl maltoside) and the mixture was incubated on ice for at least one hour. A 25 ml sample containing 500 µg 25 of the protein in the refolding buffer was applied to a Pro-RPC HR 3/5 reversed phase chromatography column. The applied sample contained 500 µg protein in 25 ml of refolding buffer. To the column was then applied a solution B comprised of 99.9% H<sub>2</sub>O + 0.1% trifluoroacetic acid (TFA). A 10 ml volume of solution C [10% H<sub>2</sub>O, 90% acetonitrile (AcN) + 0.1% TFA] was applied to the column at a 0 - 60% gradient into 30 solution B at a flow rate of 0.5ml/min. and a fraction size of 0.5ml. The fractions were monitored at A214; 2.0 absorbance units full scale (AUFS).

Fractions containing the protein (corresponding to peak 1) were 35 pooled for renaturation by stepwise dialysis. The fractions were first dialysed in 0.1% TFA in 25% glycerol overnight at 4°C ; then dialyzed in 0.01% TFA in 25% glycerol overnight at 4°C; then dialyzed in 0.001% TFA in 25% glycerol for 3.0 hours; then dialyzed for 3 hours at 4°C in 50 mM NaPO<sub>4</sub>, pH 6.0, 10 mM dithiotreitol (DTT) in 25% glycerol. The

- 28 -

protein was then dialyzed for 3.0 hours at 4°C in 50 mM NaPO<sub>4</sub>, pH 7.0, 0.15 M NaCl, 10 mM DTT in 25% glycerol; and then finally dialyzed in 50 mM NaPO<sub>4</sub>, pH 7.8, 0.3 M NaCl, 10 mM DTT, 0.2% Tween 20 in 25% glycerol. This resulted in purified, refolded, soluble, active HCV NS3

5 protease.

Far UV circular dichroism (CD) analysis of the protein was used to monitor the refolding from an acid denatured state to a folded state at neutral pH. The protein recovery was monitored by a UV scan and SDS-  
10 PAGE analysis.

**Results:**

**Detergent-assisted Refolding of His-HIV183**

15 HisHIV183 was quantitatively extracted from an *E. coli* inclusion body pellet. SDS-PAGE analysis at the various stages of extraction shows that sequential washes are essential to remove significant amounts of the contaminating proteins. HisHIV183 was extracted from the washed  
20 inclusion body pellet in the presence of 5M GuHCl. The 5M GuHCl extract was applied to a SEPHACRYL S-300 column and the appropriate fractions were pooled based on SDS-PAGE analysis. The amino acid sequence of the first ten residues was verified.

25 Refolding was performed at very low concentrations of protein, in the presence of DTT, lauryl maltoside and glycerol at 4°C. The diluted protein was concentrated on a Pro-RPC reversed phase column. Two peaks were obtained based on the UV and protein profile. Only Peak 1 has yielded soluble protein after stepwise dialysis. Far UV CD spectral  
30 analysis was used to monitor refolding from a denatured state at acid pH to a folded state at neutral pH. At pH 7.4, the protein was found to exhibit significant amounts of secondary structure that is consistent with that of beta sheet protein. At low pH, the CD spectrum showed that it is fully random coil, having a minimal molar ellipticity at 200nm. The  
35 ratio of this minimum at 200nm to that of the shoulder at 220 nm is approximately 4:1. This ratio decreased when the secondary structure formation occurred at neutral pH.

- 29 -

A UV scan at each step of dialysis showed that the protein recovery was >90% up to pH 7.4 and that there was no light scattering effect due to protein aggregates. SDS-PAGE analysis also indicated that there was no loss of protein up to pH 7.0 during refolding. Precipitation 5 of protein occurred at the last step of dialysis, and the soluble protein was clarified by centrifugation. The overall protein recovery was about 0.10%. The refolded protein was found to be active in a trans-cleavage assay using the *in vitro*-translated 5A/5B substrate in the presence of 4A peptide.

10

Example 8

Analysis of Refolded NS3 Protease Activity by  
*In Vitro* Translation Assay

15

To detect HCV NS3 protease activity in *trans*, we have expressed a 40 kD protein containing the NS5A/5B cleavage site in cell-free translation system and used that as the substrate for the enzyme. The substrate protein produces two protein products of apparent molecular 20 weight 12.5 kD (NS 5A') and 27 kD (NS5B') upon cleavage by the HCV NS3 protease.

The plasmid pTS102 encoding the substrate 5A/5B was linearized by digestion with EcoR I and was transcribed using T7 RNA polymerase 25 *in vitro*. The RNA was translated in presence of  $^{35}\text{S}$  methionine in rabbit reticulocyte lysates according to the manufacturer's (Promega ) protocol to produce HCV specific protein. In a 20  $\mu\text{l}$  total reaction mixture containing 10mM Tris, pH 7.5, 1mM DTT, 0.5mM EDTA, and 10% glycerol was placed 2 to 8  $\mu\text{l}$  of  $^{35}\text{S}$  methionine-labeled translated 30 5A/5B substrate. The reaction was started with the addition of 10 $\mu\text{l}$  of HCV NS3 protease (SEQ ID NO: 5) with an approximately equimolar amount (2  $\mu\text{M}$ ) of the carboxyterminal 33 mer cofactor NS4A (SEQ ID NO: 29) in solubilization buffer (50mM Na Phosphate, pH 7.8, 0.3M NaCl, 0.2% Tween 20, 10 mM DTT or BME, 10% glycerol), and incubated 35 at 30°C for about one hour. Reactions were stopped by adding an equal volume of 2X Laemmli sample buffer (Enprotech Inc.) and heating at 100°C for 3 minutes. Reaction products were separated by SDS PAGE electrophoresis; gels were fixed, dried and subjected to autoradiography.

The assay was able to cleave 5A/5B substrate in a dose responsive manner, producing the expected cleaved products: 5A (12.5 kD) and 5B (27 kD) as shown by SDS PAGE analysis. The production of cleaved 5A 5 and 5B polypeptides from the 5A/5B substrate is proof that soluble, active, refolded HCV protease was indeed produced by the process of example 7.

10

Example 9Surface Plasmon Resonance Assay

15 The present example illustrates a method for determining if a compound can be useful as an HCV protease inhibitor using the surface plasmon resonance assay. Figures 8A and 8B illustrate the technique.

BLAcore™ is a processing unit for Biospecific Interaction Analysis. The processing unit integrates an optical detection system 20 with an autosampler and a microfluidic system. BLAcore™ uses the optical phenomena, surface plasmon resonance to monitor interaction between biomolecules. SPR is a resonance phenomenon between incoming photons and electrons on the surface of thin metal film. Resonance occurs at a sharply defined angle of incident light. At this 25 angle, called the resonance angle, energy is transferred to the electrons in the metal film, resulting in a decreased intensity of the reflected light. SPR response depends on a change in refractive index in the close vicinity of the sensor chip surface, and is proportional to the mass of analyte bound to the surface. BLAcore continuously measure the 30 resonance angle by a relative scale of resonance units (RU) and displays it as an SPR signal in a sensorgram, where RU are plotted as a function of time.

In addition, BLAcore™ uses continuous flow technology. One 35 interactant is immobilized irreversibly on the sensor chip, comprising a non-crosslinked carboxymethylated dextran providing a hydrophilic environment for bimolecular interaction. Solution containing the other interactant flow continuously over the sensor chip surface. As

- 31 -

molecules from the solution bind to the immobilized ligand, the resonance angle changes resulting in a signal registered by the instrument.

5        In this methodology, the enzymatic reactions are carried out outside of the BIAcore, i.e. in reaction tubes or 96-well tissue culture plates, as it is conventionally done for any of the currently available high throughput assays. The SPR is only used as a detection means for determination of the amount of an intact substrate remaining in a  
10      solution with and without the enzyme after the reaction is quenched.

In order to measure the amount of the intact substrate prior to the addition of enzyme, a means of capturing the substrate onto the sensor chip had to be established. In addition, to satisfy the requirement for a  
15      high throughput assay on the BIAcore, the substrate needed to be removed from the surface subsequent to completion of analysis. This is required since the same surface will be used for the subsequent reactions. To accomplish these two requirements, a phosphotyrosine is synthetically attached to one end of the substrate. The phosphotyrosine  
20      was chosen due to the commercial availability of an anti-phosphotyrosine monoclonal antibody. The antibody is covalently attached to the sensor chip by standard amine coupling chemistry. The anti-phosphotyrosine antibody, bound permanently to the chip is used to capture the phosphotyrosine-containing substrate in a reversible  
25      manner. The antibody-phosphotyrosine interaction is ultimately used to capture and release the peptide substrate when desired by regeneration of the surface with various reagents i.e. 2 M MgCl<sub>2</sub>.

30      Introduction of the intact peptide onto the antibody surface results in a larger mass which is detected by the instrument. To follow the extent of peptide cleavage, a mixture of peptide substrate and enzyme is incubated for the desired time and then quenched. Introduction of this mixture containing the cleaved peptide and the intact peptide to a regenerated antibody surface results in a lower mass  
35      value than that detected for a sample containing only intact peptide. The difference in the two values is then used to calculate the exact amount of intact peptide remaining after cleavage by the enzyme.

Although the reduction in mass can be directly followed with many large substrates, due to the small mass of a typical synthetic peptide substrate (10-20 amino acids, 1-3 Daltons), the mass difference, and thus the signal difference between the intact and cleaved peptide is 5 very small within the signal to noise ratio of the instrument. To circumvent this low sensitivity, we attached a biotin at the N-terminus of the peptide. By addition and thus tagging of peptide with streptavidin prior to injection of tagged peptide onto the antibody surface of the chip, the signal due to the presence of streptavidin will be higher. Using this 10 approach, a cleaved peptide lacking the N-terminal half, tagged with streptavidin will result in a much lower signal.

The HCV protease 5A-5B peptide substrate, DTEDVVACSMSYTWTGK (SEQ ID NO 18) was synthesized with an 15 additional phosphotyrosine at the C-terminus and biotin at the N-terminus. The biotin was then tagged with streptavidin. An anti-phosphotyrosine monoclonal antibody, 4G10 (Upstate Biotechnology Inc., Lake Placid, New York) was coupled to the sensor chip. In the 20 absence of HCV protease, the intact, streptavidin-tagged biotinylated phosphotyrosine peptide results in a large signal (large mass unit/large signal) through its interaction with the anti-phosphotyrosine monoclonal antibody (Mab).

The protease-catalyzed hydrolysis of the phosphotyrosine-biotinylated peptide was carried out in a 96 well plate. The reaction was 25 stopped with an equal volume of mercuribenzoate. The cleaved peptide which lacks the tagged streptavidin (less mass) results in the loss of response units (lower signal).

30 Using this method, numerous compounds can be tested for their inhibitory activity since the antibody surface can be regenerated repetitively with 2 M MgCl<sub>2</sub>.

Procedure for Coupling Anti-phosphotyrosine Mab to the Sensor Chip

35

The anti-phosphotyrosine Mab is coupled to the carboxymethylated dextran surface of a sensor chip in the following manner. The flow rate used throughout the coupling procedure is 5

- 33 -

$\mu$ l/min. The surface is first activated with a 35  $\mu$ l injection of NHS/EDC (N-hydroxysuccinimide/N-dimethylaminopropyl-N'-ethylcarbodiimide-HCl). This is followed by a 40 ml injection of Mab 4G10 at 50  $\mu$ g/ml in 10 mM sodium acetate buffer, pH=4.0. Any

5 remaining activated esters are then blocked by the injection of 35  $\mu$ l of 1 M ethanolamine. These conditions result in the immobilization of approximately 7,500 response units (420  $\mu$ M) of antibody.

Binding of Peptide and Regeneration of Mab 4G10 Surface

10

The flow rate used throughout the BIACore analysis run is 5  $\mu$ l/min. A 4  $\mu$ l injection containing streptavidin-tagged peptide (peptide concentration at 2 $\mu$ M, streptavidin binding sites concentration at 9 $\mu$ M) is carried out. The amount of streptavidin-tagged peptide

15 bound to the antibody surface (in response units) is measured 30 seconds after the injection is complete.

Regeneration of sensor chip surface

20

Regeneration of the Mab 4G10 surface is achieved using a 4  $\mu$ l pulse of 2 M MgCl<sub>2</sub> after each peptide injection. Surfaces regenerated up to 500 times still showed 100% binding of tagged peptide.

Determination of the Optimal Concentration of Peptide and

25 Streptavidin

To determine the optimal peptide concentration, a standard curve was generated using various amounts of peptide (0-10  $\mu$ M) in the presence of excess streptavidin. A value in the linear range, 2  $\mu$ M, was chosen for standard assay conditions.

The amount of streptavidin required to completely tag the peptide was determined using a peptide concentration of 2.5  $\mu$ M and titrating the amount of streptavidin ( $\mu$ M of binding sites). All the

30 peptides were shown to be completely tagged when streptavidin concentrations greater than 3  $\mu$ M (approximately equimolar to the peptide concentration) were used. A streptavidin concentration of 9  $\mu$ M (a 4.5 fold excess) was chosen for standard assay conditions.

Application of Described Methodology to HCV Protease

5

The HCV protease 5A/5B peptide substrate, DTEDVVACSMSYTWTGK (SEQ ID NO 18), with phosphotyrosine at the C-terminal and biotin at the N-terminal is synthesized. Anti-phosphotyrosine monoclonal antibody, 4G10 was coupled to the sensor 10 chip.

In the absence of HCV protease, the intact streptavidin-tagged biotinylated phosphotyrosine peptide results in a large signal (large mass unit/large response units) through its interaction with the anti-phosphotyrosine monoclonal antibody. 15

The protease-catalyzed hydrolysis of the phosphotyrosine-biotinylated peptide was carried out in a 96 well plate. The reaction was stopped with an equal volume of the quenching buffer containing 20 mercuribenzoate. Streptavidin was added to tag the peptide which binds to the biotin. The cleaved peptide which lacks the tagged streptavidin (less mass) results in the loss of response units.

Using this assay, numerous compounds can be tested for their 25 inhibitory activity since the antibody surface can be regenerated repetitively with 2 M MgCl<sub>2</sub>.

The peptide cleavage activity by HCV protease can be monitored in a time dependent manner using the BIACore-based methodology. 30 Using the concentrated enzyme and the BIACore substrate, Biotin-DTEDVVAC SMSYTWTGK-pY (SEQ ID NO 17), 50% substrate cleavage is achieved within 1 hour using the BIACore-based HCV assay. Based on the amount of enzyme, His-NS3(183)Δ4AHT needed to reach a 35 50% cleavage within 2 hours, a time scale desired for a development of a high throughput assay, we estimate that 1 liter of fermentation of the His-NS3(183)Δ4AHT construct results in enough protease to run at least 100 reactions on the BIACore.

- 35 -

Standard Operating Procedure for BIACore-based HCV Assay

Reactions are prepared in a 96-well tissue culture plate using the Reaction Buffer (50 mM HEPES, pH 7.4, 20 % glycerol, 150 mM NaCl, 5 mM EDTA, 0.1% Tween-20, 1 mM DTT ) as diluent. The final reaction volume is 100  $\mu$ l. Sample with the peptide alone (Biotin-DTEDVVAC SMSYTWTGKpY) is prepared by addition of 10  $\mu$ l of peptide stock at 100  $\mu$ M (prepared in the reaction buffer) to 90  $\mu$ l of reaction buffer, so that the final concentration of peptide is 10  $\mu$ M. Samples comprised of 10 peptide and the enzyme are prepared by addition of 10  $\mu$ l of peptide stock at 100  $\mu$ M and 10  $\mu$ l of partially purified His-NS3 (183)- $\Delta$ 4A-HT stock at 1.7 mg/ml (both prepared in the reaction buffer) to 80  $\mu$ l of reaction buffer, so that the final concentration of peptide and the enzyme is 10 and 0.1  $\mu$ M respectively. The reaction is held at 30°C for 15 the specified time and then quenched. Quenching is achieved by transferring a 20- $\mu$ l aliquot of the reaction mixture to a new tissue culture plate containing an equal volume of PMB Quenching Buffer (50 mM HEPES, pH 7.8, 150 mM NaCl, 5 mM P-Hydroxymercuribenzoic Acid, and 13 mM EDTA).

20 To prepare the quenched reaction mixture for injection onto the sensor surface, 30  $\mu$ l PMB BIACore Buffer (50 mM HEPES, pH 7.4, 1 M NaCl) and 30  $\mu$ l of streptavidin at 0.5 mg/ml in water is added to the 40  $\mu$ l of the quenched reaction mixture to a final volume of 100  $\mu$ l. In this 25 step, all the peptides are tagged with streptavidin prior to the injection of samples. Finally, 4  $\mu$ l of this sample is injected over the antiphosphotyrosin surface for determination of the intact versus cleaved peptide. The final concentration of peptide and the streptavidin in the BIACore sample is 2 and 9  $\mu$ M respectively.

30

### Experimental Conditions:

5	<u>Substrate:</u> ID	Biotin-DTEDVVAC SMSYTWGK-pY (SEQ NO 19) in Reaction buffer without DTT
10	<u>Concentration:</u>	170 µM (Crude peptide, based on weight)
15	<u>Enzyme:</u>	10 µl of concentrated His-NS3 (183)-Δ4A-HT at 1.7 mg/ml
20	<u>Reaction volume:</u>	100 µl
25	<u>Reaction buffer:</u>	50 mM HEPES, pH 7.8 20 % glycerol 150 mM NaCl 1mM EDTA 1mM DTT 0.1% Tween-20
	<u>Temp:</u>	30° C
	<u>Quench with:</u>	<i>p</i> -hydroxymercuribenzoate

- 37 -

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

(i) APPLICANT: Schering Corporation

10 (ii) TITLE OF INVENTION: Hepatitis C Protease Having a Hydrophilic Motif

(iii) NUMBER OF SEQUENCES: 27

15 (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Schering Corp.
- (B) STREET: 2000 Galloping Hill Road
- (C) CITY: Kenilworth
- (D) STATE: New Jersey
- 20 (E) COUNTRY: USA
- (F) ZIP: 07033-0530

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- 25 (B) COMPUTER: Apple Macintosh
- (C) OPERATING SYSTEM: Macintosh 7.1
- (D) SOFTWARE: Microsoft Word 5.1a

(vi) CURRENT APPLICATION DATA:

- 30 (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- 35 (A) APPLICATION NUMBER: 08/440,409
- (B) FILING DATE: 12 MAY 1995

(viii) ATTORNEY/AGENT INFORMATION:

- 38 -

(A) NAME: Lunn, Paul G.  
(B) REGISTRATION NUMBER: 32,743  
(C) REFERENCE/DOCKET NUMBER: JB0494

5 (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: 908-298-5061  
(B) TELEFAX: 908-298-5388

(2) INFORMATION FOR SEQ ID NO:1:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 549 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:  
20 (A) NAME/KEY: HCV NS3 Protease

25 GCG CCC ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG 45  
Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly  
1 5 10 15

30 TGT ATA ATC ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG 90  
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu  
20 25 30

GGT GAG GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA 135  
Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala  
35 40 45

35 ACG TGC ATC AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA 180  
Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly  
50 55 60

- 39 -

ACG AGG ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT 225  
 Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr  
 65 70 75

**5**  
 ACC AAT GTG GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT 270  
 Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly  
 80 85 90

**10** TCC CGC TCA TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC 315  
 Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr  
 95 100 105

**15** CTG GTT ACG AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT 360  
 Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly  
 110 115 120

**20** GAT AGC AGG GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA 405  
 Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu  
 125 130 135

**25** AAA GCC TCC TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC 450  
 Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala  
 140 145 150

**30** GTG GGC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG 495  
 Val Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys  
 155 160 165

**35** GCG GTG GAC TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA 540  
 Ala Val Asp Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg  
 170 175 180

TCC CCG GTG  
**35** Ser Pro Val

- 40 -

## (2) INFORMATION FOR SEQ ID NO:2:

## 5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: peptide

Arg Lys Lys Lys Arg Arg

## 15 (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 567 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 25 (ix) FEATURE:

## (A) NAME/KEY:

GCG CCC ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG 45

Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly

30 1

5

10

15

TGT ATA ATC ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG 90

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu

20

25

30

35

GGT GAG GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA 135

Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala

35

40

45

- 41 -

ACG TGC ATC AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA 180  
 Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly  
 50 55 60

**5**  
 ACG AGG ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT 225  
 Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr  
 65 70 75

**10** ACC AAT GTG GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT 270  
 Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly  
 80 85 90

TCC CGC TCA TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC 315  
**15** Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr  
 95 100 105

CTG GTT ACG AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT 360  
 Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly  
**20** 110 115 120

GAT AGC AGG GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA 405  
 Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu  
 125 130 135

**25**  
 AAA GGC TCC TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC 450  
 Lys Gly Ser Ser Gly Pro Leu Leu Cys Pro Ala Gly His Ala  
 140 145 150

**30** GTG GGC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG 495  
 Val Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys  
 155 160 165

GCG GTG GAC TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA 540  
**35** Ala Val Asp Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg  
 170 175 180

- 42 -

TCC CCG GTG AGA AAG AAG AAG AGA AGA  
 Ser Pro Val Arg Lys Lys Lys Arg Arg

## (2) INFORMATION FOR SEQ ID NO:4:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 603 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

15 (A) NAME/KEY: pBJ1022(His/NS3 (182)/H.T.

20 ATG AGA GGA TCG CAT CAC CAT CAC CAC ACG GAT CCG CCC ATC 45  
 Met Arg Gly Ser His His His His His His His Thr Asp Pro Pro Ile  
 1 5 10 15

25 ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG TGT ATA ATC 90  
 Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile  
 20 25 30

30 ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG GGT GAG GTC 135  
 Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val  
 35 40 45

35 CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA ACG TGC ATC 180  
 Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile  
 50 55 60

40 AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA ACG AGG ACC 225  
 Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr  
 65 70 75

- 43 -

ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT GTG 270  
 Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val

80 85 90

5 GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT TCC CGC TCA 315  
 Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser  
 95 100 105

TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC CTG GTT ACG 360  
 10 Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr  
 110 115 120

AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT GAT AGC AGG 405  
 Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg  
 15 125 130 135

GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA AAA GGC TCC 450  
 Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser  
 140 145 150

20 TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC GTG GGC CTA 495  
 Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu  
 155 160 165

25 TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG GCG GTG GAC 540  
 Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys Ala Val Asp  
 170 175 180

TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA TCC CCG GTG 585  
 30 Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg Ser Pro Val  
 185 190 195

AGA AAG AAG AAG AGA AGA  
 Arg Lys Lys Lys Arg Arg

35

- 44 -

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 630 base pairs

5 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

10

## (ix) FEATURE:

(A) NAME/KEY: pT5His/HIV/183 No solubilizing motif

	ATG	AGA	GGA	TCG	CAT	CAC	CAT	CAC	CAT	CAC	GGA	TCC	CAT	AAG	GCA	45
15	Met	Arg	Gly	Ser	His	Gly	Ser	His	Lys	Ala						
	1				5				10					15		
	AGA	GTT	TTG	GCT	GAA	GCA	ATG	AGC	CAT	GGT	ACC	ATG	GCG	CCC	ATC	90
20	Arg	Val	Leu	Ala	Glu	Ala	Met	Ser	His	Gly	Thr	Met	Ala	Pro	Ile	
		20				25								30		
	ACG	GCG	TAC	GCC	CAG	CAG	ACG	AGA	GGC	CTC	CTA	GGG	TGT	ATA	ATC	135
25	Thr	Ala	Tyr	Ala	Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	
		35				40								45		
	ACC	AGC	CTG	ACT	GGC	CGG	GAC	AAA	AAC	CAA	GTG	GAG	GGT	GAG	GTC	180
30	Thr	Ser	Leu	Thr	Gly	Arg	Asp	Lys	Asn	Gln	Val	Glu	Gly	Glu	Val	
		50			55									60		
	CAG	ATC	GTG	TCA	ACT	GCT	ACC	CAA	ACC	TTC	CTG	GCA	ACG	TGC	ATC	225
35	Gln	Ile	Val	Ser	Thr	Ala	Thr	Gln	Thr	Phe	Leu	Ala	Thr	Cys	Ile	
		65			70									75		
	AAT	GGG	GTA	TGC	TGG	ACT	GTC	TAC	CAC	GGG	GCC	GGA	ACG	AGG	ACC	270
40	Asn	Gly	Val	Cys	Trp	Thr	Val	Tyr	His	Gly	Ala	Gly	Thr	Arg	Thr	
		80			85									90		

- 45 -

	ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT GTG	315
	Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val	
	95	100
		105
5	GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT TCC CGC TCA	360
	Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser	
	110	115
		120
10	TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC CTG GTT ACG	405
	Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr	
	125	130
		135
15	AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT GAT AGC AGG	450
	Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg	
	140	145
		150
20	GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA AAA GGC TCC	495
	Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser	
	155	160
		165
25	TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC GTG GGC CTA	540
	Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu	
	170	175
		180
30	TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG GCG GTG GAC	585
	Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys Ala Val Asp	
	185	190
		195
35	TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA TCC CCG GTG	630
	Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg Ser Pro Val	
	200	205
		210

**(2) INFORMATION FOR SEQ ID NO:6:**

### **(i) SEQUENCE CHARACTERISTICS:**

5 (A) LENGTH: 162 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

**10 (ii) MOLECULE TYPE: cDNA**

**(ix) FEATURE:**

(A) NAME/KEY: NS4A

15 AGC ACC TGG GTG CTC GTT GGC .GGC GTC CTG GCT GCT CTG GCC GCG 45  
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala  
1 5 10 15

20 TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC 90  
 Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val  
 20 25 30

25 TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC 135  
 Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr  
 35 40 45

CAG GAG TTC GAT GAG ATG GAA GAG TGC 162  
Gln Glu Phe Asp Glu Met Glu Glu Cys  
50

**(2) INFORMATION FOR SEQ ID NO:7:**

### **(i) SEQUENCE CHARACTERISTICS:**

### (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 47 -

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

## (A) NAME/KEY: NS3 +NS4A

5

GGC CCC ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG 45  
 Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly  
 1 5 10 15

10

TGT ATA ATC ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG 90  
 Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu  
 20 25 30

15

GGT GAG GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA 135  
 Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala  
 35 40 45

20

ACG TGC ATC AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA 180  
 Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly  
 50 55 60

25

ACG AGG ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT 225  
 Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr  
 65 70 75

30

ACC AAT GTG GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT 270  
 Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly  
 80 85 90

35

TCC CGC TCA TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC 315  
 Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr  
 95 100 105

40

CTG GTT ACG AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT 360  
 Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly  
 110 115 120

- 48 -

GAT AGC AGG GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA 405  
 Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu  
 5 125 130 135

AAA GGC TCC TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC 450  
 Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala  
 10 140 145 150

GTG GCC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG 495  
 Val Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys  
 155 160 165

15 GCG GTG GAC TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA 540  
 Ala Val Asp Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg  
 170 175 180

TCC CCG GGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG 585  
 20 Ser Pro Gly Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala  
 185 190 195

TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC 630  
 Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val  
 25 200 205 210

TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC 675  
 Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr  
 215 220 225

30 CAG GAG TTC GAT GAG ATG GAA GAG TGC 702  
 Gln Glu Phe Asp Glu Met Glu Glu Cys  
 230

- 49 -

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 855 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

10

## (ix) FEATURE:

(A) NAME/KEY: pNB182Δ4AHT

15 ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC CAT AAG GCA 45  
Met Arg Gly Ser His His His His His Gly Ser His Lys Ala

1 5 10 15

AGA GTT TTG GCT GAA GCA ATG AGC CAT GGT ACC ATG GCG CCC ATC 90  
20 Arg Val Leu Ala Glu Ala Met Ser His Gly Thr Met Ala Pro Ile

20 25 30

ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG TGT ATA ATC 135  
25 Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile

35 40 45

ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG GGT GAG GTC 180  
30 Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val

50 55 60

CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA ACG TGC ATC 225  
35 Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile

65 70 75

35 AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA ACG AGG ACC 270  
Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr

80 85 90

- 50 -

ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT GTG 315  
 Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val  
 95 100 105

5  
 GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT TCC CGC TCA 360  
 Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser  
 110 115 120

10 TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC CTG GTT ACG 405  
 Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr  
 125 130 135

15 AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT GAT AGC AGG 450  
 Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg  
 140 145 150

20 GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA AAA GGC TCC 495  
 Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser  
 155 160 165

25 TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC GTG GGC CTA 540  
 Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu  
 170 175 180

30 TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG GCG GTG GAC 585  
 Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys Ala Val Asp  
 185 190 195

35 TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA TCC CCG GGG 630  
 Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg Ser Pro Gly  
 200 205 210

35 GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG TAT TGC CTG 720  
 Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu  
 215 220 225

- 51 -

TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC TTG TCC GGG 765  
Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly  
230 235 240

5 AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAG TTC 810  
Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe  
245 250 255

GAT GAG ATG GAA GAG TGC CGG AAG AAA AAG AGA CGC AAG CTT AAT 855  
10 Asp Glu Met Glu Glu Cys Arg Lys Lys Lys Arg Arg Lys Leu Asn  
260

15

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 711 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

25

## (ix) FEATURE:

## (A) NAME/KEY:

GGC CCC ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA CGG 45  
Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly  
30 1 5 10 15

TGT ATA ATC ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG 90  
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu  
20 25 30

35

GGT GAG GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA 135  
Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala  
35 40 45

- 52 -

ACG TGC ATC AAT GGG GTA TGC TGG ACT GTC TAC CAC CGG GCC GGA 180  
 Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly  
 50 55 60

**5**  
 ACG AGG ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT 225  
 Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr  
 65 70 75

**10** ACC AAT GTG GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT 270  
 Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly  
 80 85 90

TCC CGC TCA TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC 315  
**15** Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr  
 95 100 105

CTG GTT ACG AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT 360  
 Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly  
**20** 110 115 120

GAT AGC AGG CGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA 405  
 Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu  
 125 130 135

**25**  
 AAA GGC TCC TCG CGG CGT CCG CTG TTG TGC CCC GCG GGA CAC GCC 450  
 Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala  
 140 145 150

**30** GTG GGC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG 495  
 Val Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys  
 155 160 165

GCG GTG GAC TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA 540  
**35** Ala Val Asp Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg  
 170 175 180

TCC CCG GGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG 585

Ser Pro Gly Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala

185

190

195

5

TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC 630

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val

200

205

210

10

TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC 675

Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr

215

220

225

CAG GAG TTC GAT GAG ATG GAA GAG AAG GAG ACA GAG

15

Gln Glu Phe Asp Glu Met Glu Glu Lys Glu Thr Glu

230

**(2) INFORMATION FOR SEQ ID NO:10:**

20

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 855 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

**(ii) MOLECULE TYPE: cDNA****(ix) FEATURE:****(A) NAME/KEY:**

30

ATG AGA GGA TCG CAT CAC CAT CAC CAC ACG GAT CCG GCG CCC

Met Arg Gly Ser His His His His His His Thr Asp Pro Ala Pro

1

5

10

15

35

ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA CGG TGT ATA 45

Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile

20

25

30

- 54 -

ATC ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG GGT GAG 90  
 Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu

35 40 45

5 GTC CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA ACG TGC 135  
 Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys  
 50 55 60

10 ATC AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA ACG AGG 180  
 Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg  
 65 70 75

15 ACC ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT 225  
 Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn  
 80 85 90

GTG GAC CAA GAC CTT GTG GGC TGG CCC GCT CCT CAA GGT TCC CGC 270  
 20 Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg  
 95 100 105

TCA TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC CTG GTT 315  
 Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val  
 25 110 115 120

ACG AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT GAT AGC 360  
 Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser  
 30 125 130 135

AGG GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA AAA GGC 405  
 Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly  
 35 140 145 150

- 55 -

	TCC TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC GTG GGC 450		
	Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly		
5	155	160	165
	CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG GCG GTG 495		
	Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys Ala Val		
	170	175	180
10	GAC TTT ATC CCT GTG GAG AAC CTA GAC ACA ACC ATG AGA TCC CCG 540		
	Asp Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg Ser Pro		
	185	190	195
	GGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG TAT TGC 585		
15	Gly Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys		
	200	205	210
	CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC TTG TCC 630		
	Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser		
20	215	220	225
	GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAG 675		
	Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu		
	230	235	240
25	TTC GAT GAG ATG GAA GAG AAG GAG ACA GAG 705		
	Phe Asp Glu Met Glu Glu Lys Glu Thr Glu		
	245	250	
30			

## (2) INFORMATION FOR SEQ ID NO:11:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

- 56 -

(D) TOPOLOGY: double

(ii) MOLECULE TYPE: cDNA

5       GA TCA   CCG GTC TAG ATCT  
          T   GGC CAG ATC TAGA

(2) INFORMATION FOR SEQ ID NO:12:

10       (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY:

20

CCG   GTC   CGG   AAG   AAA   AAG   AGA CGC TAG   C  
      AG   GCC   TTC   TTT   TTC   TCT GCG ATC   G

(2) INFORMATION FOR SEQ ID NO:13:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 79 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

35       (A) NAME/KEY:

CCG GCA   ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAA TTC  
      GT   TAA TAT GGA CTG TCC CTC CAA GAG ATG GTC CTT AAG

- 57 -

GAT GAG ATG GAA GAG TGC CGG AAG AAA AAG AGA CGC A  
CTA CTC TAC CTT CTC ACG GCC TTC TTT TTC TCT GCG TTC GA

**(2) INFORMATION FOR SEQ ID NO:14:**

5

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear

**(ii) MOLECULE TYPE:** polypeptide

**(ix) FEATURE:**

15 (A) NAME/KEY: NS4A Active Mutant

Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys  
5 10

**20 (2) INFORMATION FOR SEQ ID NO:15:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- 25 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

**(ii) MOLECULE TYPE:** polypeptide

30 **(ix) FEATURE:**

(A) NAME/KEY: NS4A Active Mutant

Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys  
5 10

35

**(2) INFORMATION FOR SEQ ID NO:16:**

**(i) SEQUENCE CHARACTERISTICS:**

- 58 -

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

**(ii) MOLECULE TYPE: polypeptide**

### **(ix) FEATURE:**

**(A) NAME/KEY: Soluble 5A/5B Substrate**

10

Asp Thr Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp Thr  
5 10 15  
Gly Lys

15

**(2) INFORMATION FOR SEQ ID NO:17:**

**(i) SEQUENCE CHARACTERISTICS:**

20

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

**(ii) MOLECULE TYPE: polypeptide**

**(ix) FEATURE:**

**(A) NAME/KEY: Mutant 5A/5B Substrate**

30

Asp Thr Glu Asp Val Val Ala Cys Ser Met Ser Tyr Thr Trp Thr  
5 10 15  
Gly

INFORMATION FOR SEQ ID NO:182

35

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 amino acids  
(B) TYPE: amino acid

- 59 -

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: polypeptide

5

(ix) FEATURE:

(A) NAME/KEY: Mutant Soluble 5A/5B Substrate

Asp Thr Glu Asp Val Val Ala Cys Ser Met Ser Tyr Thr Trp Thr

10

5

10

15

Gly Lys

2) INFORMATION FOR SEQ ID NO:19:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: polypeptide

(ix) FEATURE:

(A) NAME/KEY: Soluble 5A/5B Substrate

25

Asp Thr Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp Thr

5

10

15

Gly Lys Tyr

2) INFORMATION FOR SEQ ID NO:20:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: polypeptide

- 60 -

(ix) FEATURE:

(A) NAME/KEY: Soluble 5A/5B Substrate

5 Asp Thr Glu Asp Val Val Ala Cys Ser Met Ser Tyr Thr Trp Thr  
5 10 15  
Gly Lys Tyr

2) INFORMATION FOR SEQ ID NO:21:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: polypeptide

(ix) FEATURE:

20 (A) NAME/KEY: Soluble 4B/5A Substrate

Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu  
5 10 15  
25 Arg Asp Ile Trp Asp

2) INFORMATION FOR SEQ ID NO:22:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: polypeptide

(ix) FEATURE:

(A) NAME/KEY: histidine tag

Met Arg Gly Ser His His His His His His Thr Asp Pro  
5 10

5

## 2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: polypeptide

15

## (ix) FEATURE:

(A) NAME/KEY: hydrophilic tail

20 Arg Lys Lys Lys Arg Arg Lys Leu Asn

5

## 2) INFORMATION FOR SEQ ID NO:24:

25

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30

## (ii) MOLECULE TYPE: polypeptide

## (ix) FEATURE:

(A) NAME/KEY: hydrophilic tail

35

Lys Glu Thr Glu

## 2) INFORMATION FOR SEQ ID NO:25:

- 62 -

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: polypeptide

## 10 (ix) FEATURE:

- (A) NAME/KEY: hydrophilic tail

Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu  
15 5 10 15

Arg Asp Ile Trp Asp  
20

## 20 (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 base pairs
- (B) TYPE: nucleic acid
- 25 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 30 (ix) FEATURE:

- (A) NAME/KEY: NS4A Mutant

GTG CTC GTT GGC GGC GTC CTG GCT CTG GCT CTG GCC GCG TAT TGC CTG 45  
Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu  
35 1 5 10 15

TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC TTG TCC GGG 90  
Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly

- 63 -

20

25

30

AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAG TTC 135  
 Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe  
 5 35 40 45

GAT GAG ATG GAA GAG TGC  
 Asp Glu Met Glu Glu Cys

50

10

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 810 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

20

## (ix) FEATURE:

- (A) NAME/KEY: pNB182Δ4AHT

25 ATG AGA GGA TCG CAT CAC CAT CAC CAC ACG GAT CCG CCC ATC 45  
 Met Arg Gly Ser His His His His His His His Thr Asp Pro Pro Ile  
 1 5 10 15

ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG TGT ATA ATC 90  
 30 Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile  
 20 25 30

ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG GGT GAG GTC 135  
 Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val  
 35 35 40 45

- 64 -

CAG ATC GTG TCA ACT GCT ACC CAA ACC TTC CTG GCA ACG TGC ATC 180  
Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile  
50 55 60

5 AAT GGG GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA ACG AGG ACC 225  
Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr  
65 70 75

10 ATC GCA TCA CCC AAG GGT CCT GTC ATC CAG ATG TAT ACC AAT GTG 270  
Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val  
80 85 90

15 GAC CAA GAC CTT GTG GCC TGG CCC GCT CCT CAA GGT TCC CGC TCA 315  
Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser  
95 100 105

20 TTG ACA CCC TGC ACC TGC GGC TCC TCG GAC CTT TAC CTG GTT ACG 360  
Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr  
110 115 120

25 AGG CAC GCC GAC GTC ATT CCC GTG CGC CGG CGA GGT GAT AGC AGG 405  
Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg  
125 130 135

30 GGT AGC CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA AAA GGC TCC 450  
Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser  
140 145 150

35 TCG GGG GGT CCG CTG TTG TGC CCC GCG GGA CAC GCC GTG GGC CTA 495  
Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu  
155 160 165

- 65 -

TTC AGG GCC GCG GTG TGC ACC CGT GGA GTG ACC AAG GCG GTG GAC 540  
Phe Arg Ala Ala Val Cys Thr Arg Gly Val Thr Lys Ala Val Asp  
5 170 175 180

TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC ATG AGA TCC CCG GGG 585  
Phe Ile Pro Val Glu Asn Leu Glu Thr Thr Met Arg Ser Pro Gly  
10 185 190 195

GTC CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG TAT TGC CTG 630  
Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu  
15 200 205 210

TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC TTG TCC GGG 720  
Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly  
20 215 220 225

AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC CAG GAG TTC 765  
Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe  
25 230 235 240

25 GAT GAG ATG GAA GAG TGC CGG AAG AAA AAG AGA CGC AAG CTT AAT 810  
Asp Glu Met Glu Glu Cys Arg Lys Lys Lys Arg Arg Lys Leu Asn  
245 250 255

## 30 (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 162 base pairs  
(B) TYPE: nucleic acid  
35 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

- 66 -

## (ix) FEATURE:

## (A) NAME/KEY: Native NS4A

TCA ACA TGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG 45

5 Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala  
1 5 10 15

TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT GTC 90

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val  
10 20 25 30

TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC 135

Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr  
35 40 45

15

CAG GAG TTC GAT GAG ATG GAA GAG TGC

Gln Glu Phe Asp Glu Met Glu Glu Cys

50

20

## (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 162 base pairs

25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

30

## (ix) FEATURE:

## (A) NAME/KEY: Native NS4A

TCA ACA TGG GTG CTC GTT GGC GGC GTC CTG GCT GCT CTG GCC GCG 45

Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala

35 1 5 10 15

- 67 -

TAT TGC CTG TCA ACA GCA TGC GTG GTC ATA GTG GGC AGG ATT GTC 90

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val

20

25

30

5 TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT CTC TAC 135

Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr

35

40

45

CAG GAG TTC GAT GAG ATG GAA GAG TGC

10 Gln Glu Phe Asp Glu Met Glu Glu Cys

50

**(2) INFORMATION FOR SEQ ID NO:29:**

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acid residues

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: polypeptide

(ix) FEATURE:

(A) NAME/KEY: Carboxyl 33 mer of NS4A

25

Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala

5

10

15

Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met

30

20

25

30

Glu Glu Cys

**35 (2) INFORMATION FOR SEQ ID NO:30:**

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acid residues

- 68 -

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: polypeptide

(ix) FEATURE:

- (A) NAME/KEY: Carboxl 33 mer of NS4A of HCV-BK strain

10 Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala  
5 10 15

Ile Val Pro Asp Arg Glu Leu Leu Tyr Gln Glu Phe Asp Glu Met  
20 25 30

15

Glu Glu Cys

20

25

- 69 -

WE CLAIM:

WE CLAIM:

- 5    1. A soluble HCV NS3 protease which is bacterially produced, denatured and refolded to produce soluble, active HCV NS3 protease.
- 10    2. A soluble HCV protease comprising HCV NS3 protease fused to a solubilizing motif.
- 15    3. The soluble HCV protease of claim 2 wherein the solubilizing motif is a polypeptide comprising ionizable amino acids which, when fused to the HCV NS3 protease produces a protein which is soluble when expressed in *E. coli*.
- 20    4. The soluble HCV protease of claim 3 wherein the solubilizing motif is comprised of Lys or Arg amino acid residues.
- 25    5. The soluble HCV protease of claim 2 wherein the solubilizing motif is defined by SEQ ID NO: 2, SEQ ID NO: 23 or SEQ ID NO: 24.
- 30    6. The soluble HCV protease of claim 2 having a sequence defined by SEQ ID NO 3 or SEQ ID NO 4.
- 35    7. An HCV protease comprising HCV NS3 protease fused to an HCV NS4A cofactor wherein said cofactor has been modified by deletion or substitution of one or more amino acid residues to prevent cleavage by the HCV NS3 protease.
- 30    8. The HCV protease of claim 7 further comprising a solubilizing motif fused to said protease.
- 35    9. The soluble HCV protease of claim 8 wherein the solubilizing motif is a polypeptide comprising a sequence of ionizable amino acids which, when fused to said HCV protease produces a protein that is soluble when expressed in *E. coli*.

- 70 -

10. The soluble HCV protease of claim 9 wherein the sequence of ionizable amino acids comprises Lys or Arg amino acid residues.
- 5 11. The soluble HCV protease of claim 9 wherein the sequence of ionizable amino acids is defined by SEQ ID NO: 2, SEQ ID NO: 23 or SEQ ID NO: 24.
12. The HCV protease of claim 8 having a sequence defined by SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10 or SEQ ID NO: 27.
- 10 13. A modified NS4A cofactor having a sequence defined by SEQ ID NO 14, SEQ ID NO 15 or SEQ ID NO 26.
- 15 14. An isolated nucleic acid or a vector encoding a polypeptide or protein of claims 1-13.
15. A host cell transformed or transfected with an isolated nucleic acid or vector of claim 14.
- 20 16. The host cell of claim 15 wherein the host cell is bacterial.
- 25 17. A method for making a soluble HCV protease comprising culturing the host cell of claims 15 or 16 under conditions in which the protease is expressed.
- 25 18. An HCV NS3 protease having a polyhistidine tag attached to said protease, wherein the histidine tag is comprised of two or more histidine residues.

30

35

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : <b>C12N 9/50, 15/51, 15/57, 15/62, C12Q 1/37</b></p>		<p>A3</p>	<p>(11) International Publication Number: <b>WO 96/36702</b> (43) International Publication Date: 21 November 1996 (21.11.96)</p>
<p>(21) International Application Number: <b>PCT/US96/06387</b> (22) International Filing Date: <b>9 May 1996 (09.05.96)</b> (30) Priority Data: 08/440,409 12 May 1995 (12.05.95) US</p>		<p>(81) Designated States: AL, AM, AU, AZ, BB, BG, BR, BY, CA, CN, CZ, EE, GE, HU, IS, JP, KG, KR, KZ, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p>	
<p>(71) Applicant: SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033 (US). (72) Inventors: DASMAHAPATRA, Bimalendu; 322 Grant Avenue, Nutley, NJ 07110 (US). MURRAY, Michael, G.; 8 Fawn Drive, Lebanon, NJ 08833 (US). RAMANATHAN, Lata; 32 Valley Way, West Orange, NJ 07052 (US). BUTKIEWICZ, Nancy, J.; 1000 Oak Lane, Plainfield, NJ 07060 (US). (74) Agents: LUNN, Paul, G. et al.; Schering-Plough Corporation, Patent Dept. K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US).</p>		<p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p> <p>(88) Date of publication of the international search report: 16 January 1997 (16.01.97)</p>	
<p>(54) Title: SOLUBLE, ACTIVE HEPATITIS C VIRUS PROTEASE</p> <p>(57) Abstract</p> <p>Soluble HCV NS3 protease, including the NS3 protease fused to a solubilizing motif. A fusion of the NS3 and NS4 regions under conditions where they are not cleaved by the NS3 protease. Bacterially expressed soluble HCV NS3 protease. Host cells wherein at least 1 % of the cell's total protein is soluble HCV NS3 protease.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

**INTERNATIONAL SEARCH REPORT**

Int'l Application No  
PCT/US 96/06387

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 C12N9/50 C12N15/51 C12N15/57 C12N15/62 C12Q1/37

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>JOURNAL OF VIROLOGY, vol. 67, no. 7, July 1993, pages 3835-3844, XP002010751  BARTENSCHLAGER, R. ET AL.: "Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions" see "Expression of HCV fusion proteins and generation of antisera" at page 3836, right column  see figure 1</p> <p>---</p> <p>-/-</p>	1-6, 14-18

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

1

Date of the actual completion of the international search

Date of mailing of the international search report

13 August 1996

03.12.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentstaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+ 31-70) 340-3016

Authorized officer

Hillenbrand, G

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 96/06387

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>JOURNAL OF VIROLOGY, vol. 67, no. 7, July 1993, pages 4017-4026, XP000561255 TOMEI, L. ET AL.: "NS3 is a serine protease required for processing of hepatitis C virus polyprotein" see page 4018, under "Induction of expression plasmids and preparation of fusion proteins" ---</p>	1-6, 14-17
A	<p>GENE, vol. 67, 1988, pages 31-40, XP002010752 SMITH, D.B. AND JOHNSON, K.S.: "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase" see the whole document -----</p>	1

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US 96/06387

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
  
3.  Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. claims 1-6 and 13-18
2. claims 7-12 and 14-17
3. claims 13-17

\* see continuation-sheet PCT/ISA/210 \*

1.  As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3.  As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-6 and 13-18

**Remark on Protest**

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

**INTERNATIONAL SEARCH REPORT**

International Application No. PCT/US 96/ 06387

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/210**

1. Soluble HCV NS3 protease fused to a solubilizing motif, an isolated nucleic acid or vector encoding such a protease, a (bacterial) host cell transformed or transfected with such an isolated DNA or vector, a method for making such a soluble HCV protease comprising culturing such a host and an HCV NS3 protease having a polyhistidine tag.
2. HCV protease comprising HCV NS3 protease fused to an NS4A cofactor which has been modified to prevent cleavage by the HCV NS3 protease (see claim 7), such a polypeptide further comprising a solubilizing motif fused thereto, an isolated nucleic acid or vector encoding such a polypeptide, a (bacterial) host cell transformed or transfected with such an isolated DNA or vector, and a method for making such a (soluble) polypeptide comprising culturing such a host cell.
3. Modified NS4A cofactor, an isolated nucleic acid or vector encoding such a cofactor, a (bacterial) host cell transformed or transfected with such as isolated DNA or vector, and a method for making such a cofactor comprising culturing such a host cell.